METHODEN IN DER EVALUATION

Zeitschrift für Evaluation | 24. Jahrgang | 2025
Heft 2 | S. 245–263 | https://doi.org/10.31244/zfe.2025.02.04 |
Waxmann CC BY 4.0

Qualitative Comparative Analysis as a Method for Evaluation Studies

Felix Hörisch[1], *Kai Brumm*[2]

Abstract: Qualitative Comparative Analysis (QCA) is an increasingly popular method in evaluation studies. QCA aims at systematic, formalised comparisons of cases on the basis of set theory. It uses Boolean algebra to detect necessary and sufficient conditions for an outcome of interest. The article discusses the advantages and drawbacks of the method in the context of evaluations and explains the core concepts of QCA – both for its Crisp- and Fuzzy-Set variants. Moreover, it presents key steps of the QCA procedure when applied to evaluation studies, including calibration, the search for necessary conditions, the setup of truth tables, and different solution types. Building on this, we briefly elaborate on preconditions under which QCA may be a valuable tool for evaluation research.

Keywords: Qualitative Comparative Analysis (QCA), evaluation, method, necessary condition, sufficient condition, set theory.

Qualitative Comparative Analysis als Methode für Evaluationsstudien

Zusammenfassung: Qualitative Comparative Analysis (QCA) ist eine zunehmend beliebte Methode in Evaluationsstudien. QCA zielt auf systematische, formalisierte Vergleiche von Fällen mithilfe von Mengenlehre. Dabei wird Boolesche Algebra genutzt, um notwendige und hinreichende Bedingungen für ein interessierendes Outcome zu ermitteln. Der Artikel erörtert Vorzüge und Nachteile der Methode im Kontext von Evaluationen und erläutert die Kernkonzepte von QCA – sowohl für ihre Crisp- wie auch ihre Fuzzy-Set-Variante. Ferner beschreibt er die wichtigsten analytischen Schritte, die bei der Anwendung der Methode auf Evaluationsstudien zu beachten sind, einschließlich der Kalibrierung, der Fahndung nach notwendigen Bedingungen, der Konstruktion von Wahrheitstafeln sowie der verschiedenen QCA-Lösungsarten. Abschließend werden knapp die Kernvoraussetzungen, unter denen QCA ein wertvolles Instrument für die Evaluationsforschung darstellen kann, zusammengefasst.

Schlagwörter: Qualitative Comparative Analysis (QCA), Evaluation, Methode, notwendige Bedingung, hinreichende Bedingung, Mengentheorie.

1 University of Applied Labour Studies (HdBA), Federal Employment Agency, Mannheim, Germany.
2 Institute of Political Science, Heidelberg University, Heidelberg, Germany; School of Social Sciences, University of Applied Sciences HTW Saar, Saarbrücken, Germany.

1. Introduction

Qualitative Comparative Analysis (QCA) is an increasingly popular method in the social sciences that has been developed since the late 1980s. Its foundations were laid out in three seminal monographs by Charles Ragin, an American sociologist (Ragin, 1987, 2000, 2008). Throughout the last decades, QCA left its mark on various disciplines, also enriching the methodological toolkit of evaluation research (Gerrits & Verweij, 2016; Thiem, 2014, p. 489).[3]

So, what characterises Qualitative Comparative Analysis and when is it used? Essentially, the term refers to both a research approach and a technical method of analysis. As an approach, QCA aims at formalised inspections of set relations. This is done by systematically comparing (a medium number of) empirical cases based on set theory.[4] So, the methodological approach is recommended if social reality can and shall be modelled in terms of subset and superset relations. Furthermore, as a technical method, QCA seeks to identify necessary and sufficient conditions for an outcome of interest by means of Boolean algebra (Mello, 2021; Oana et al., 2021).[5] Given this dual character, it is not surprising that the literature faces a hard time categorising QCA (Wagemann, 2023): Sometimes, it is portrayed as a "middle way" between quantitative and qualitative techniques (Ragin, 1987), while other authors classify QCA as a qualitative method (e. g., Blatter et al., 2018). In line with Wagemann (2023), we describe QCA as a valuable instrument for comparative case studies that capitalises on some strengths of systematic quantitative methods. For instance, it can easily deal with a medium number of cases (e. g., 15 to 30 units of observation) due to its systematic approach and its mathematical foundations (Hörisch & Heiken, 2020). Nonetheless, the number of cases (*N*) should never be the sole motivation for conducting a Qualitative Comparative Analysis. Rather, such a justification for applying QCA is weak and falls far short the mark.

Let us elaborate on this and clarify under what conditions QCA is an appropriate method that may deliver worthwhile insights. In fact, the set-theoretic toolkit comes along with a number of specific assumptions about social reality. These assumptions can best be illustrated by comparing QCA to the classic linear regression model. Let us suppose we assembled panel data on youth unemployment in European countries throughout the last decades. Then, we used linear regression and found that youth unemployment is affected by the business cycle, the existence of a dual system of vocational training, and public spending on active labour market policies (ALMPs). So,

3 The website of "COMPASSS" (*Comparative Methods for Systematic Cross-Case Analysis*), a worldwide network of QCA scholars, lists numerous studies applying QCA. Some of them can be attributed to evaluation research. See *https://compasss.org/bibliography*.

4 In practice, many research designs relying on QCA strive to explain a *moderate* number of cases emphasising in-depth case knowledge as an important feature of the QCA approach. However, the set-theoretic method *per se* allows for pure macro (or *large-N*) analyses as well. These do not aim at understanding individual cases. So, set theory is a defining feature of QCA, while the explanation of single cases is not. We thank an anonymous reviewer for pointing this out.

5 Put simply, the QCA vocabulary turns dependent variables into 'outcomes' and independent ones into 'conditions'. However, it is justified to use different terms in a QCA context, as our article will show.

based on the regression equation, we were to conclude that public expenditure for certain labour market programmes had a negative effect on unemployment among young persons, ceteris paribus. The nature of this inference is twofold. First, it refers to variables, not to specific cases (European countries). More precisely, it describes the linear effect of a single variable on youth unemployment, provided that other covariates are held constant. Second, such statements are symmetric: The more a country spends on ALMPs, the lower youth unemployment will be. Conversely: The less is spent, the more the youth unemployment rate will increase.

By contrast, QCA aims at gaining insights both into the (complex) causation of social phenomena and into specific cases. Scholars applying this method do not seek to deduct *generalisable* statements about the relationship between different variables. Instead, for example, QCA users would ask why youth unemployment is low in countries 1, 2 and 3; they try to fully identify the reasons for this outcome in each case. To answer their question, they would first define what low youth unemployment is, specifying a maximum unemployment rate that successful countries will undercut. Based on this definition, scholars could calibrate the set of nations having 'low' youth unemployment (Y). The same would have to be done when calibrating the set of EU member states witnessing economic growth (G) and the set of countries having a dual system of vocational training (D). Finally, researchers would decide which cases belong to the set of nations that spend 'a lot of money' on active labour market programmes (A) and which do not. In consequence, a hypothetical QCA result could look like this: $G^*A + D \rightarrow Y$. Please note that QCA employs Boolean algebra which is based on logical operators. So, our hypothetical solution would entail two paths leading to the outcome: Economic growth AND high public expenses for active labour market policies OR the existence of a dual system of vocational training both represent sufficient conditions for low youth unemployment.

What are the properties of this result? Referring to it, we can easily illustrate three assumptions of causal complexity that form the basis of every Qualitative Comparative Analysis (Schneider & Wagemann, 2012, pp. 76–90). First, our hypothetical example shows that many roads may lead to Rome. While the interplay of economic growth and generous ALMPs may contain youth unemployment in country 1, dual training systems ensure the occurrence of the outcome in countries 2 and 3. We speak of equifinality to denote that various conditions might imply one and the same outcome. So, the latter has several mutually non-exclusive explanations. Second, some conditions might only have an impact if they occur together with others (*conjunctural causation*; Hörisch, 2012). We can understand this by looking at country 1, where a booming economy alone is not enough to bring about the outcome. Instead, tailored active labour market programmes are needed there to guarantee low numbers of unemployed among the younger generation. Third, set relations are not symmetric but asymmetric: Stating that D is a sufficient condition for Y says nothing about the outcome if condition D is absent. So, if a country does not have a dual system of vocational training, it might encounter high levels of unemployment among the young. However, we cannot draw this conclusion by only relating to a QCA for outcome Y. Instead, countries without dual vocational training schemes may well benefit from other conditions limiting joblessness in younger age cohorts (Oana et al., 2021, p. 8).

Moreover, causal complexity and case orientation hint at another important feature of the method. Those who use Qualitative Comparative Analysis are interested in the causes of effects, not the effects of certain causes (Oana et al., 2021, pp. 4–9). One should choose QCA if one strives to illuminate the reasons why a certain phenomenon such as low youth unemployment occurs – why it occurs in some cases but not in others (Mahoney & Goertz, 2006). The set-theoretic method is not suitable when gauging the impact of specific variables and its magnitude. To summarise, researchers should employ QCA if they seek to fathom certain phenomena and expect complex causal relationships that can be modelled by means of set theory. Compared to that, the question whether the number of cases is medium or not turns secondary. Nonetheless, one's knowledge of individual cases, which is important for well-founded QCAs, is likely to suffer if too many cases are considered.

So, its methodological foundations distinguish QCA both from case studies and classic statistical methods such as regression analyses. In comparison to comparative case studies (for a path-breaking overview, see Gerring, 2008), QCA is characterised by a more formalised approach that enables scholars to process a higher number of cases. Compared to regressions, QCA assumes causal complexity, is far more case-oriented and qualifies for causes-of-effects research questions. Irrespective of this, like with any method in evaluation studies, it is decisive that the analysis relies on careful theoretical reasoning, a thoroughly crafted research design, and valid data (Hörisch & Heiken, 2020; Oana et al., 2021). Qualitative Comparative Analyses may be combined with single-case studies to cross-validate their findings and to unveil the causal mechanisms behind social phenomena (Rohlfing & Schneider, 2018).

The remainder of this method note is structured as follows: The next section elaborates on the preconditions under which QCA is a valuable tool for evaluation research. Building on this, the third section briefly explains core concepts of any QCA and discusses the main analytical and technical steps that must be managed when applying the method. The fourth section concludes.

2. When is QCA an Appropriate Method for Evaluation Studies?

Compared to other techniques, Qualitative Comparative Analysis is still a young method, especially in the realm of evaluation research. Nevertheless, it has evolved substantially throughout the last years, established itself in the social sciences and has gained some recognition among evaluation scholars (Befani, 2013; Gerrits & Verweij, 2016; Thiem, 2014). For instance, earlier studies employed QCA to evaluate German education systems (Freitag & Schlicht, 2009), vaccination programmes (Glatman-Freedman et al., 2010), irrigation projects (Lam & Ostrom, 2010) or national evaluation societies (Holvoet & Dewachter, 2013). Its growing popularity revealed drawbacks and advantages of QCA. On the one hand, standard procedures as to how analytical problems should be dealt with, are only just emerging. This includes the challenge of limited empirical diversity, which is a common problem both for evaluation studies and for social science research in general. At least, QCA as an approach makes this challenge explicit with QCA theorists passionately discussing how to cope with theo-

retically conceivable but unobserved cases (Schneider, 2016; Thomann, 2017; Duşa, 2022; Baumgartner & Thiem, 2020; Thiem, 2022). On the other hand, by differentiating between necessary and sufficient conditions, for instance, QCA may enable scholars to break new ground and tackle gaps in evaluation research.

Typically, QCA is portrayed as a case-oriented approach. Each case that is to be examined must be judged individually as to what extent it belongs to a set of interest. This could be the set of countries having a dual system of vocational training, for example. Thus, QCA requires a higher level of case knowledge than popular quantitative methods such as linear regression (Blatter et al., 2007, pp. 216–217). Moreover, the set-theoretic method is suitable if one builds on strong theoretical expectations and elaborated hypotheses in the sense of causal complexity. A good QCA checks for both necessary and sufficient conditions that might be associated with an outcome of interest (Lauth et al., 2009, p. 135). Likewise, it examines the effects of negated conditions (e. g., non-dual systems of vocational education) and causes of the negated outcome (i. e., non-low youth unemployment), thereby enriching the methodological toolkit of evaluation research (Jäckle, 2015, p. 204; Hörisch & Heiken, 2020). Specifically, QCA is capable of illuminating different paths to success or non-success of programmes under evaluation, e. g., regarding labour market policies or development cooperation. Furthermore, the technique can handle medium numbers of cases although this fact alone should not motivate its use. However, this is relevant for evaluators, who often have to limit their analyses to a few cases. If one intends to compare – let us say – 16 federal states in Germany, 19 legislative periods, 24 schools in a district or 14 regions within a country, these rather small case numbers need not discourage researchers from using QCA (Jäckle, 2015, p. 204; Lauth et al., 2009, p. 119). Perhaps, one of the greatest advantages of QCA for evaluation studies is the concept of equifinality. Equifinality means that different conditions or combinations of conditions may lead to the same outcome. This implies that individual conditions – or their negations – can occur in various explanatory paths that lead to the outcome under consideration.

In contrast to these advantages, the drawbacks accompanying the set-theoretic method must not be ignored. Notably, some scholars argue that QCA does not qualify for *large-N* studies, because such research designs are likely to prevent the evaluator from getting in-depth knowledge of individual cases. According to this view, though, it is precisely this case knowledge that is essential for a 'good' QCA (see Vis, 2012, p. 170; Rutten, 2022; Finn, 2022). On the other hand, QCA is not that useful for evaluation designs analysing single-digit numbers of cases. The calibration stage of a QCA entails categorising individual cases, which results in a loss of information. The latter is particularly severe when small samples are examined. Moreover, limited empirical diversity (Ragin & Sonnett, 2005; Thiem, 2022) is likely to become a significant issue if only a few cases are studied, while several conditions are suspected to cause the outcome. So, given a (very) small number of cases, traditional qualitative methods of comparison should be applied (Gerring, 2008).

Besides, QCA solution terms may turn out to be complicated comprising several combinations of conditions that imply the outcome. Thus, it can be challenging to communicate clear results to the addressees of an evaluation report when the analysis builds on the set-theoretic method. QCA findings are often not concise and catchy, but complex and convoluted. This is all the more true, the more conditions are includ-

ed and the lower empirical diversity among the cases becomes (Hörisch & Heiken, 2020). Accordingly, some QCA scholars recommend choosing a number of conditions that is reasonably compatible with the number of cases (Lauth et al., 2009, p. 127).

This article concentrates on the most common variants of the set-theoretic method, namely Crisp-Set QCA (csQCA) and Fuzzy-Set QCA (fsQCA) (Ragin, 2000). The latter – allowing for gradual set membership scores between 0 and 1 – is frequently used in evaluation studies. Further varieties of Qualitative Comparative Analysis include Multi-Value QCA (mvQCA) that relies on set memberships like 2, 3, or 4 in addition to the conventional values 0 and 1.[6] Besides, temporal QCA (tQCA; Schneider & Wagemann, 2012, pp. 263–274) attempts to address the problem of temporality, i. e. the sequence of conditions and their outcomes. However, both variants are not very widespread yet. Due to their early stage of development, the use of mvQCA and tQCA is controversial (Hörisch & Heiken, 2020).

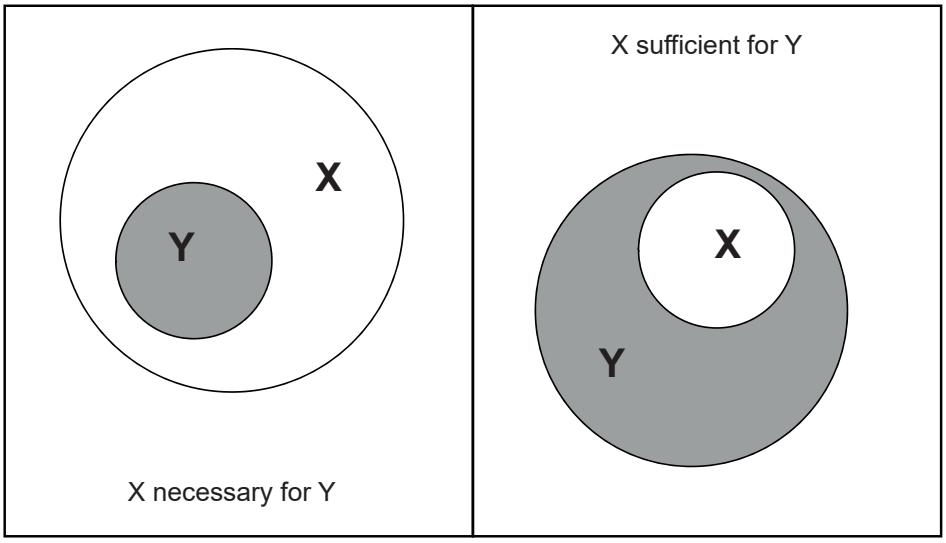
3. Qualitative Comparative Analysis in a Nutshell

This section discusses core steps of Qualitative Comparative Analysis thereby defining important concepts and *parameters of fit* on which the method relies. As mentioned in the introduction, QCA seeks to express relationships between sets in terms of necessary and sufficient conditions. Using Venn diagrams (Fig. 1), we can visualise these ideal set relations: On the one hand, a condition X is called necessary for outcome Y if the condition occurs in all cases in which Y is present. Adding another hypothetical set to our example, this means that in *all* countries marked by *low youth unemployment* (Y) the condition *powerful trade unions* (U) would be present if we diagnosed that U was necessary for Y. So, by implication, there should be no instances where the outcome occurs, but U is missing. The condition is a superset of Y, it includes the outcome because all cases showing Y are members of the condition set as well. On the other hand, we detect a sufficient condition if its occurrence always implies the outcome. For instance, if all countries having a dual system of vocational training (D) also show *low youth unemployment* (Y), then D can be seen as sufficient for the outcome. Thus, D represents a subset of Y because all cases within D also exhibit Y. Please note that we generally denote conditions with an X when we are not discussing hypothetical examples (see Figures 1–3).

To identify necessary and sufficient conditions based on a sample of cases, researchers must first calibrate the sets of interest. Calibration is the essential prerequisite for any QCA. It entails several steps with users precisely analysing their cases drawing on theoretical and empirical knowledge. Then, they must decide whether or to what extent the cases belong to a particular set. To express these assessments, researchers assign membership scores to each case regarding any condition and the outcome. If these scores are dichotomous with zeros indicating non-members, while ones mark (perfect) set membership, we speak of crisp sets. For example, if we only distinguish between countries having (value of 1) or not having (value 0) a dual system of

6 Berg-Schlosser and Cronqvist (2012, pp. 169–200) as well as Schneider and Wagemann (2012, pp. 255–262) offer lucid introductions to mvQCA.

Figure 1: Venn diagrams for a necessary (left) and a sufficient condition (right).



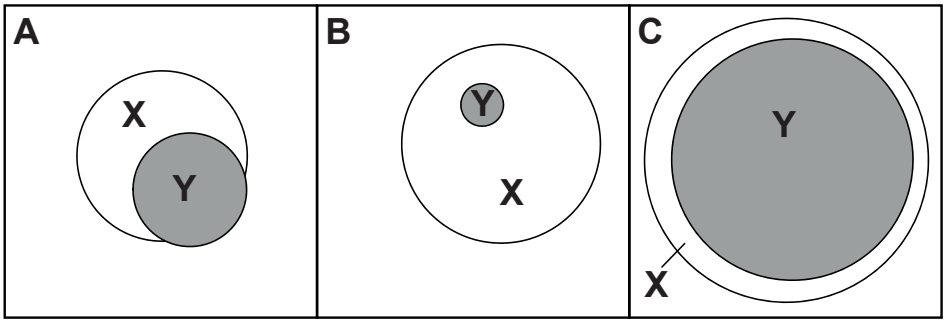
Source: Own depiction

vocational training (D), we are constructing a crisp set. By contrast, fuzzy sets allow for gradual membership scores between 0 and 1 such as 0.9 or 0.33. That is, fuzzy sets can express the different *degrees* to which cases belong to a set, not just the fundamental differences between members and non-members (Buche & Siewert, 2015, pp. 387–389). Here, we could distinguish between nations counting on an *encompassing* dual system of vocational training (1), countries with dual training schemes in many but not all pertinent job sectors (0.7), and so on. Depending on the underlying theoretical concept, researchers often have the choice of converting empirical information either into binary crisp or more fine-grained fuzzy sets. Concretely, dichotomous variables must lead to crisp sets, whereas variables having multiple categories or metric variables can be transposed into both fuzzy and crisp sets. Please note that assigning fuzzy membership scores of 0.5 can cause analytical problems as this basically signifies that a case is neither a member nor a non-member of a particular set. Therefore, the literature typically recommends avoiding neutral fuzzy scores of 0.5 (Rubinson et al., 2019, p. 3; Schneider & Wagemann, 2012, p. 28; Buche & Siewert, 2015, p. 393).[7]

7 It is not possible to assign cases with set membership scores of 0.5 to a specific truth table row, see below. Nevertheless, such 'indifferent' cases are not dropped from the sample but virtually stay in the background: If calibration reflects case knowledge and there remains good reason why one cannot decide whether a case is more in the set, or more out of the set, the assignment of 0.5 can be the decision to make. We thank an anonymous referee for emphasising that such indifferences regarding single conditions or the outcome can technically be managed. Precisely, cases that show membership scores of 0.5 do influence all consistency and coverage scores being computed during the QCA procedure (see below). However, as pointed out, they do not show up in the truth table because the QCA algorithm assigns each case to one truth table row and one row only. This is not possible for indifferent cases having a 0.5 set membership. Thus, they are not a member of any configuration in the truth table.

There are several procedures to calibrate sets based on empirical data. However, such procedures primarily concern the mathematical and technical implementation of this QCA stage. Irrespective of this, a good QCA is built on carefully defined concepts that guide the calibration. Having theoretical concepts in mind, one should specify qualitative anchors (or threshold values) that will assign the cases to different set membership scores. This is done by locating the cases of the respective sample between the anchors. In other words, it is not important which mathematical function one uses for calibration or whether a case has a fuzzy score of 0.75 or 0.76. Instead, it is essential that the calibration relies on a solid conceptual foundation and links set memberships back to theoretical reasoning. Apart from this principle, there are three remarks on calibration that we would like to discuss here. First, most types of raw data (categorial or metric variables) may be transposed either into crisp or fuzzy sets – both options are available. The only exceptions are variables being naturally dichotomous, such as participation in an election: Either a citizen has cast her vote or not. So, she is either a member of the (crisp) set 'voters' or she is a non-member in case of voting abstention; it is not reasonable to calibrate a fuzzy set here. Second, it might be challenging to identify theoretically meaningful thresholds that mark substantial differences between cases. This is especially true for numerical raw variables such as youth unemployment: Which anchor or crossover point could separate countries with 'low' youth unemployment from nations having a 'large' share of jobless young people? Moreover, while crisp sets require a single qualitative anchor, fuzzy sets involve at least three threshold values for full membership, perfect non-membership, and as a crossover point. The latter establishes a qualitative difference *in kind* between cases (Oana et al., 2021, pp. 32–34). If there is a lack of theoretical arguments that could justify one's qualitative anchors, one may also look at how the cases are distributed over the range of the raw variables. If one observes empirical gaps in the distribution, it can be advisable to group cases according to them. Conversely, one should avoid just taking the arithmetic mean, the median or similar statistical indicators as the threshold values for calibration. This approach bears the risk of falling into mechanistic recoding without theoretical substance (Rubinson et al., 2019, p. 4). Third, it might not always be recommendable to assign set membership scores manually. If one examines a large number of cases, while having metric variables, scholars perhaps want to use software performing the 'direct' or 'indirect' methods of calibration (Ragin, 2008, pp. 85–105). Both routines yield very fine-grained fuzzy scores. To use the direct method, one again needs three qualitative anchors for exclusion, the crossover point, and complete inclusion, respectively. On this basis, the direct method employs a logistic function to transform the raw variable into set membership scores (Schneider & Wagemann, 2012, pp. 35–38). Instead, the indirect method requires the researcher to roughly organise the cases into four or six groups, for example, laying down the respective fuzzy values. Then, these preliminary fuzzy scores are regressed on the raw data using a fractional polynomial function, for instance (Duşa, 2018, pp. 92–94). Finally, the predicted values of this regression model serve as the fuzzy membership scores for all cases. Nonetheless, it should be noted that it is the qualitative anchors (or the researcher's rough categorisation of cases) which decisively determine the results of the calibration procedure. In comparison, the functions and regression models used by the direct and indirect methods are far less relevant. Instead, their fine-grained fuzzy

Figure 2: Set relations implying low *parameters of fit* for necessary conditions.



Note: Graph A illustrates consistency that is smaller than 1; Graph B shows low coverage; Graph C depicts a low relevance of necessity (RoN) score, whereas coverage in this scenario would be high.
Source: Own depiction

values might even misleadingly suggest a degree of precision that is not tenable at all if we consider the raw data. Thus, choosing theoretically substantiated anchors is the most important task when QCA practitioners calibrate sets.

When all sets of interest are calibrated, researchers can move on and search for necessary conditions.[8] To avoid analytical pitfalls, the methodological literature predominantly recommends examining necessary and sufficient conditions in two separate steps starting with necessary conditions (Schneider & Wagemann, 2012, p. 232).[9] To judge if a set is a relevant necessary condition for the outcome, QCA produces three *parameters of fit*, with each ranging between 0 and 1: consistency (or inclusion), coverage and the *relevance of necessity* (RoN) score. The three Venn diagrams in Fig. 2 illustrate set relations that imply low *parameters of fit*. Low values cast doubt on whether a meaningful necessary condition is present.

Let us first discuss consistency – a measure that reflects the quality of set relations. As mentioned above, a condition X is deemed necessary for outcome Y if Y does not occur in the absence of X. So, the outcome must be a subset of X. However, as Fig. 2A illustrates, some part of Y lies outside the condition now. Thus, there must be some cases contradicting the statement of necessity because they show the outcome without being members of set X. We can think of countries having low youth unemployment (Y) without being characterised by strong trade unions (~U). Such cases would un-

8 Some theories might imply that two or more conditions form functional equivalents when causing the outcome. Thus, at least one of them must be given for the outcome to occur. If one takes such a theory as a basis, QCA can assess whether a union of sets (two or more conditions connected by a logical OR) is necessary for Y (Schneider & Wagemann, 2012, pp. 74–75).

9 In contrast to the mainstream of QCA methodologists (Schneider & Wagemann, 2012), some authors such as Alrik Thiem advise to not dissociate "necessity relations from QCA's 'analytical moment'" (Thiem, 2014, p. 495), i. e., the construction and logical minimization of truth tables. Truth tables form the core instruments when searching for *sufficient* conditions (see below). According to Thiem, "the concept of necessity pervades the analysis of sufficiency" (2014, p. 496) which is why an upstream and isolated examination of (simple) necessary conditions might entail that researchers overlook intriguing set relations. Notwithstanding this, we continue by discussing the standard QCA procedure as laid out in established textbooks (Oana et al., 2021).

dermine the diagnosis of powerful unions being a necessary prerequisite for low levels of inoccupation among the young. As a consequence of such deviant observations, the consistency score declines below 1. Figuratively speaking, the measure corresponds to the proportion of Y that is located within the condition. Using fuzzy sets, the parameter becomes the lower, the more cases deviate from the statement of necessity and the stronger they deviate. The methodological literature suggests to no longer speak of a necessary condition if its consistency score falls below 0.9 (Schneider & Wagemann, 2012, p. 143).

Now, we may imagine an applied QCA that finds a perfect necessary condition for the outcome having a consistency score of 1. In this case, one should have a thorough look at the next two *parameters of fit*, coverage and the RoN score (see Schneider & Wagemann, 2012): If one or both of these indicators are low, this points to trivial, irrelevant necessary conditions. High values, by contrast, suggest empirically important necessary conditions (Oana et al., 2021, p. 72). Let us first touch upon coverage: Suppose you were to investigate why members of the German Bundestag vote against their party line. So, the outcome of your study would be deviating voting behaviour in parliament. Using a QCA, you could reveal that MPs only deviate from their party line if they were born in Germany. Of course, this is a meaningless, irrelevant necessary condition, since simply many more MPs were born in the Federal Republic compared to the number of dissenters in each Bundestag vote. Thus, coverage of this necessary condition would be low; the coverage score represents the size ratio between the outcome set and the necessary condition (Buche & Siewert, 2015, p. 396; see Figure 2B).

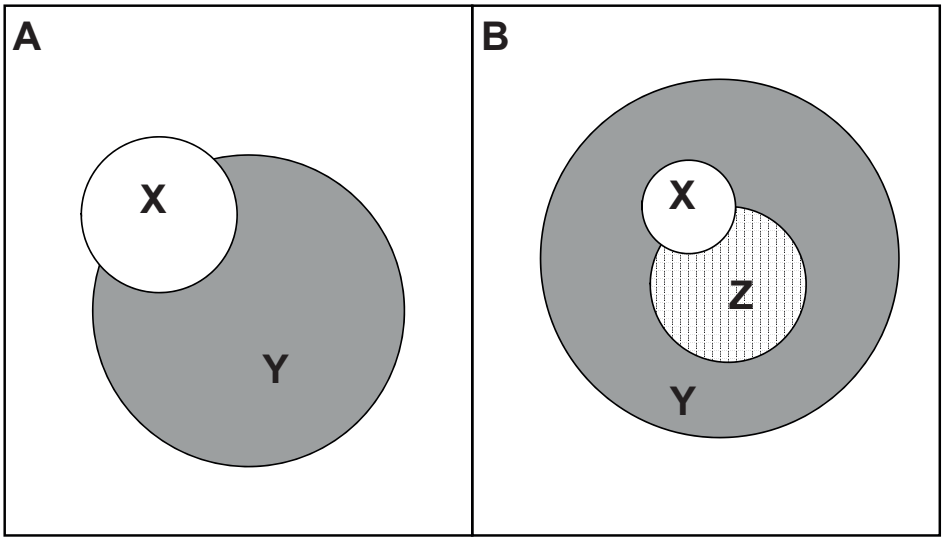
Finally, let us assume a situation in which both consistency and coverage reach acceptable levels. Nevertheless, the coverage measure only reflects the size difference between the outcome and condition X (Oana et al., 2021, p. 73). Therefore, unfortunately, one's QCA would still not be immune to detecting trivial necessary conditions because the so-called RoN value may be small (see Figure 3C). It scores low if most cases under consideration are members of the condition that is deemed to be necessary.[10] When sets are skewed in such a way, one barely observes cases in which X is absent. To say that such a condition is necessary for the outcome is therefore not too meaningful, as X is almost omnipresent or a quasi-constant, whereas its negation, ~X, barely occurs. Consequently, assuming fuzzy sets, X is likely to be a sufficiently consistent superset of *any* outcome. This probably includes both Y and its negated set, ~Y. To infer that X represents a necessary condition for Y and ~Y, though, amounts to a non-helpful, irrelevant statement (Oana et al., 2021, p. 73). Thus, the RoN measure is designed to unveil another form of triviality. We should only classify conditions as necessary if all three *parameters of fit* reach high values (Duşa, 2018, pp. 105–118).[11]

After one checked for necessary conditions, the typical QCA procedure continues by examining sufficient conditions for the outcome. Again, three *parameters of fit* are available to judge the quality and importance of a sufficient condition. As you remember, X is considered sufficient for outcome Y if X represents a subset of Y. Put

10 Strictly speaking, the RoN score is able to detect both sources of trivialness, "a big difference in size between X and Y and between X and ~X" (Oana et al., 2021, p. 73).

11 There are no commonly accepted lower boundaries for coverage necessity and the RoN score. However, from our point of view, 0.5 or 0.6 seem to be reasonable minimum thresholds.

Figure 3: Set relations implying low *parameters of fit* for sufficient conditions.



Note: Graph A visualises consistency that is smaller than 1; Graph B displays condition X having both small raw coverage and small unique coverage.
Source: Own depiction

differently, in case of a sufficient condition, X never occurs without Y; the set of X is not outside but fully inside the outcome (see Figure 1). X ensures that Y materialises.[12]

Similar to necessary conditions, the *parameters of fit* for the analysis of sufficiency range between 0 and 1 with higher values being better. The Venn diagrams in Fig. 3 demonstrate what makes two of them – consistency and coverage – decline. Typically, the methodological literature acknowledges a sufficient condition if the consistency score exceeds 0.75 or 0.8 (Schneider & Wagemann, 2012, p. 129). The parameter measures the quality of the subset relationships. As we have just established, cases that exhibit X but not outcome Y contradict the diagnosis of a sufficient condition. To exemplify this, we can imagine nations that rely on dual systems of vocational training (D) but do still witness high numbers of jobless young people (~Y). Thus, such countries would make us doubt that dual education schemes are sufficient to ensure low youth unemployment. In Figure 3A, deviant cases are located within X, the condition, but not inside the outcome. If too many observations fall into this area, consistency decreases to such an extent that we can no longer speak of a sufficient condition. So, ultimately, consistency equals the proportion of condition X that lies *inside* Y.

Coverage, in turn, is a measure that indicates the empirical relevance of the sufficient condition. Since X represents a subset of Y, it typically does not explain all cas-

12 Please note for the analysis of sufficient conditions: Here, the term 'condition' (or its placeholder X) may stand for any configuration of conditions. On the one hand, we might unveil AND combinations or intersections of conditions which are sufficient for the outcome (*conjunctural causation*). On the other hand, two or more of these configurations each can imply the outcome (*equifinality*). Such configurations occur frequently in applied QCA, be it in the course or as the findings of the analysis (Brumm, 2022). This is why you should consider X as a placeholder.

es in which the outcome occurs. For instance, generous spending on active labour market policies will probably not account for all cases in which youth unemployment is low. So, usually, there are several sufficient conditions, each of which causes Y in the sense of equifinality. Thus, a single condition becomes more important the more observations of the outcome it induces. This is the logic behind the coverage score, which basically reflects the size ratio between condition X and outcome Y. In the Venn diagram (Figure 3B), for example, condition X would not be too relevant, as it covers fewer cases compared to Z thus 'explaining' smaller parts of the outcome. What we have described so far is often referred to as *raw* coverage. If several sufficient conditions are involved, *unique* coverage is another parameter worth considering. It captures which part of the intersection of X and Y is exclusively covered by X and no other sufficient condition. In Figure 3B, conditions X and Z slightly overlap, i. e. have a small intersection. This means there are cases which exhibit both conditions. To calculate the unique coverage of X, one would take its raw coverage – figuratively speaking, the entire circle of X – and subtract the intersection with all other sufficient conditions from it – in this case the intersection with Z. Thus, unique coverage measures the extent to which a condition can claim exclusive explanatory power for the outcome (Duşa, 2018, pp. 136–138).

Apart from consistency and coverage, there is a third *parameter of fit* being relevant when QCA builds on fuzzy sets: the PRI score ('proportional reduction in inconsistency'). Again, this indicator varies between 0 and 1 and helps solving the problem of simultaneous subset relations: If X is a very small set – that is, most cases show low membership scores regarding X – then there is the risk of X being a (consistent enough) subset of both Y and its negation ~Y. This is a logical contradiction that must be avoided.[13] If simultaneous subset relations occur, it holds that we observe a *deviant case consistency in kind* (DCK) at any rate, either for the subset relation between X and Y or between X and the negated outcome (Oana et al., 2021, p. 96). Such cases contradict the statement of sufficiency as they embody the condition (set membership > 0.5), while not showing the outcome (membership score < 0.5). So, if one has detected a consistent sufficient condition, the PRI score makes one uncover simultaneous subset relationships. Having a low PRI level below 0.5 for the condition, this means that we should not classify X as sufficient for the outcome. This is because the condition simultaneously shows a high consistency score for the negation of Y and, in view of the PRI values, is more likely to be sufficient for ~Y.[14] Aside from the PRI indicator, it is important to keep an eye on true logically contradictory cases (DCKs). They also help to decide whether X must be considered a sufficient condition for Y or for its negation (Schneider & Wagemann, 2012, pp. 237–244).

Armed with this basic knowledge, we can now move on to constructing and examining truth tables. Truth tables are the most important analytical tool on which the search for sufficient conditions relies. It absorbs the empirical information provided by the cases. This information is then simplified – more formally, QCA researchers say the truth table is logically minimised. Finally, there is a solution term that is as lean as

13 Obviously, this problem requires *partial* set membership of cases. It may occur when working with fuzzy sets but cannot arise when QCA scholars use crisp sets.

14 The PRI score is available for both X → Y and X → ~Y. Thus, it is symmetric fulfilling the equation $PRI_y = 1 - PRI_{\sim y}$ (Oana et al., 2021, p. 96).

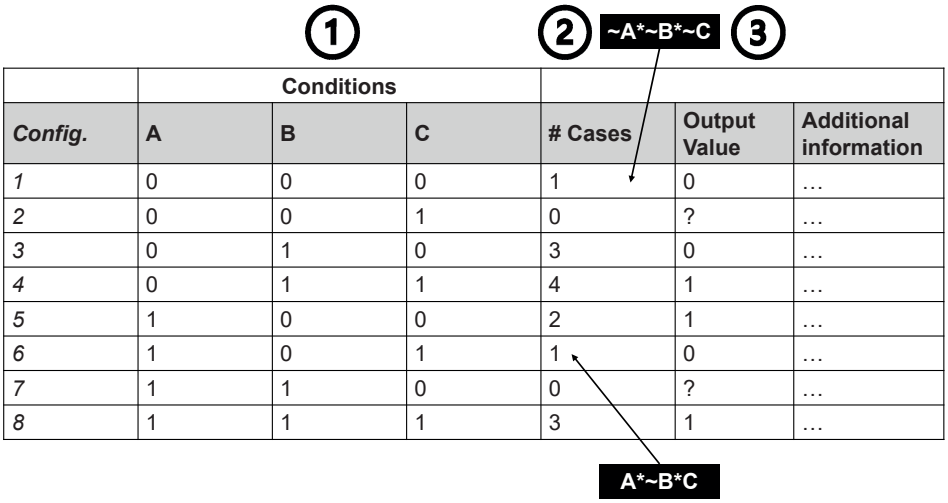
possible describing sufficient conditions for the outcome. To construct a truth table, let us assume a QCA based on (dichotomous) crisp sets. Crisp sets can be transposed directly into truth tables because each case either is a member of a given condition, or it is not. The truth table, for its part, lists all possible combinations of conditions. Each condition may be present or absent. That is, if one examines *k* conditions, the truth table will consist of *2k* rows or configurations. Each row stands for one thinkable combination of conditions (see Figure 4). As you can guess: '1' in the truth table means that the condition is given, '0' stands for its absence. In sum, the first step of our search for sufficient conditions entails setting up the truth table by combining the presence and absence of each condition and writing the configurations one below the other. Second, one must consider the crisp membership scores of all cases and assign them to the corresponding truth table row.[15] An additional column where QCA researchers fill in the (number of) cases facilitates this. However, it is likely to happen that one does not observe any cases for some configurations. The respective interplay of conditions is then theoretically conceivable but could not be observed empirically. We refer to such combinations of conditions as *logical remainders*. They characterise most applied QCAs (Brumm & Hörisch, 2018) and are an expression of limited empirical diversity. Third, one must determine the output value of each truth table row. This evaluates whether the cases of each configuration are members of outcome Y or not. Using crisp sets, one has four options. If all cases in a row show the outcome, the configuration in question can be considered sufficient for Y. One diagnoses a perfectly consistent subset relationship and assigns the output value 1. Conversely, rows in which all cases are non-members of Y are given the output value 0. In contrast, one cannot assess logical remainders as no cases with the corresponding combination of conditions occur. Remainders are usually labelled with a question mark in the truth table. Lastly, contradictory configurations, which in some cases are associated with the outcome and in other cases with ~Y, are difficult to judge. Typically, QCA practitioners resolve such contradictions by means of the raw consistency score. It must exceed certain thresholds such as 0.8 or 0.75 to presume a sufficient combination of conditions – irrespective of a (small) number of logically contradictory cases.[16]

As soon as the truth table is set up, QCA strives to logically simplify the information contained in it. Truth tables always include a few lines that have been assigned the output value 1. This means that we consider these configurations to be sufficient

15 In a sense, the configurations in a truth table represent ideal-type cases where each condition is either given or fully absent. Using fuzzy sets, QCA scholars must bear in mind that each case is partially a member of each configuration; their membership scores range between 0 and 1. However, if the fuzzy score 0.5 is consistently avoided during calibration, it applies that, given a certain case, *only one* membership score regarding the truth table rows is greater than 0.5. Thus, the case is allotted to this configuration because it comes closest to the associated ideal-type combination of conditions (Duşa, 2018, pp. 146–151).

16 When assigning output values, one should always weigh up both raw consistency and the occurrence of logically contradictory cases to decide if a certain configuration is treated as sufficient for Y or not (Rubinson et al., 2019). If one's QCA builds on crisp sets, a truth table row that shows imperfect consistency (raw consistency smaller than 1) automatically implies that there is at least one logically contradictory case. By contrast, using fuzzy sets, the consistency score of a given configuration may range below 1 without us observing a true contradiction, i. e., a case that represents the truth table row, while not showing the outcome (Schneider & Wagemann, 2012, p. 129).

Figure 4: Constructing a (hypothetical) truth table based on three conditions.



	Conditions					
Config.	**A**	**B**	**C**	**# Cases**	**Output Value**	**Additional information**
1	0	0	0	1	0	…
2	0	0	1	0	?	…
3	0	1	0	3	0	…
4	0	1	1	4	1	…
5	1	0	0	2	1	…
6	1	0	1	1	0	…
7	1	1	0	0	?	…
8	1	1	1	3	1	…

Note: The graph visualises the step-by-step construction of a truth table. 1: Listing all possible combinations of conditions. 2: Assigning cases to these configurations. 3: Specifying an output value for each truth table row. "Additional information" may include raw consistency and PRI scores, case names, etc.
Source: Own depiction

for the outcome. Accordingly, we could now write out all sufficient configurations to arrive at a solution term. However, this solution would be quite complex and difficult to interpret. Therefore, the last stage of a QCA revolves around the logical minimisation of truth tables. We want to express sufficient conditions for the outcome as succinctly as possible. Nonetheless, the solution term of a QCA looks simpler, but it actually is a logical equivalent to the original truth table rows that are included in the minimisation procedure.

When it comes to minimisation, two questions may guide our discussion: First, *how* are the configurations minimised? Second, *what* is included in the minimisation process? – Which rows of the truth table are taken into account?

The first question is answered immediately, as QCA builds on the so-called Quine-McCluskey algorithm. Originally, the algorithm was developed for circuits in electrical engineering and was eventually imported into the social sciences (Ragin, 1987). The first stage of the procedure is a systematic comparison: One compiles a list of all possible pairs of configurations that shall be minimised. This enables QCA researchers to compare each truth table row to any other. If one finds a pair that is exactly the same except for a single literal (condition), one can eliminate this (logically irrelevant) literal to arrive at a simpler expression. If this step does not allow for any further simplifications, one has found the so-called *prime implicants* (PIs). The empirically observed configurations that show output value 1 in the truth table are also called *primitive expressions*. The second step of the algorithm involves constructing a matrix that relates the prime implicants to the primitive expressions. This *prime implicants chart* is then used to eliminate all PIs that are not required to completely cover the original configurations. It should be noted that several, logically equivalent solution terms may emerge from this second stage of the algorithm (Duşa, 2018, pp. 159–160).

The second question is more interesting. Which truth table rows shall be considered when doing the logical minimisation? Depending on the answer, QCA produces different solution types. Concretely, Ragin's *Standard Analysis* (Ragin, 2008) yields a conservative (or complex) solution, a (most) parsimonious one, and an intermediate solution. All solution types make use of the empirically observed configurations that are deemed to be sufficient for the outcome (output value 1 in the truth table). If the number of these configurations is low, the solution term tends to remain complex. By contrast, combinations of conditions that are linked to the non-occurrence of the outcome (output value 0) are never subjected to the minimisation algorithm.[17] So, the solution types differ in the handling of logical remainders. In the sense of counterfactual reasoning, social scientists may draw conclusions about cases that have not been observed. One asks: What if we did? Specifically, logical remainders are combinations of conditions that did not occur among the cases analysed. If they are considered in the logical minimisation, we assume that the configurations would have been sufficient for the outcome if they had materialised. By including remainders, QCA solution terms can often be simplified further.

So, which logical remainders are used when calculating the different solution types? First, the conservative solution foregoes including any logical remainders. This implies that all remainders would not have been sufficient for the outcome if they had been observed (Buche & Siewert, 2015, p. 395). Second, the (most) parsimonious solution employs all logical remainders. They are included in the algorithm if they help to simplify the solution term. In this case, however, some remainders will probably not contribute to a more streamlined solution. Therefore, a distinction is made between simplifying assumptions – i. e., remainders that actually render the QCA solution less complex – and all other remainders that do not help accomplish this goal. However, if simplifying assumptions are used, they may include remainders that contradict our theoretical argument or knowledge. Thus, one can filter out simplifying assumptions based on theoretical expectations. QCA researchers name these theoretical presumptions '*directional expectations*'. So, a third approach is to take only those remainders into account that correspond to our theoretical arguments. This makes them '*easy counterfactuals*' and the intermediate solution uses logical remainders provided that they are easy counterfactuals. By contrast, rudiments that are not in line with theoretical knowledge are excluded as *difficult counterfactuals*. Table 1 summarises our comparison of the three solution types – also in terms of complexity and subset relations.

The so-called *Standard Analysis* ends at this point yielding the three solution types above. It is up to the researcher which solution he or she uses as a basis for interpreting QCA results. However, one should provide good reasons in favour of a certain solution type, as there is a fierce debate in the methodological literature about which one is the most appropriate (e. g., Schneider, 2016; Thomann, 2017; Duşa, 2022; Baumgartner & Thiem, 2020).

17 Thinking further, one can object that truth table rows which are marked as insufficient for the outcome (output value 0) are actually needed to determine whether a condition is a redundant element of a conjunction or a non-redundant (causally relevant) element. This cannot be decided if we consider the sufficient rows (output value 1) in isolation. See Thiem (2022, pp. 532–533) for a discussion (indirectly) exemplifying this.

Table 1: QCA solution types.

Solution	Positive configurations (output value 1)	Logical remainders	Complexity	Set relations among each other
Conservative (CS)	Yes.	None.	•••	Subset of IS und PS.
Intermediate (IS)	Yes.	*Easy counterfactuals* only.	••	Subset of PS.
Parsimonious (PS)	Yes.	All *simplifying assumptions*.	•	Superset of IS und CS.

Source: Own overview

Notwithstanding this, anyone who performs the *Standard Analysis* is by no means immune to serious analytical problems (Schneider & Wagemann, 2012). First, the problem of simultaneous subset relationships (see above) can occur with all three solution types, especially with the conservative one. Second, *untenable assumptions* might strain a QCA. These are linked to logical remainders, i. e., they may concern the most parsimonious and the intermediate solution. A distinction is made between three untenable assumptions: Some are just based on naturally impossible combinations of conditions such as 'drought' coinciding with 'flood'. Configurations like these should not be considered. The same applies to remainders that, from a technical point of view, may simplify the solutions for both the outcome and its negation. Such problematic simplifications should be avoided by removing the so-called *contradictory simplifying assumptions* from the analysis. Finally, one can fall prey to untenable assumptions if one has previously identified one or more necessary conditions. Set theory may easily show that the negation of a necessary condition can never be sufficient for one and the same outcome Y. For this reason, remainders that contain the negated necessary condition should never be included in the minimisation algorithm (Duşa, 2018, pp. 178–193). However, before diving too deep into methodological details, we conclude by briefly summarising the main characteristics of QCA as a method for evaluation research.

4. Conclusion

When drafting research designs, social scientists are required to carefully weigh up the advantages and weaknesses of different methods. Evaluation studies are no exception. This article underlined that Qualitative Comparative Analysis may enrich the methodological toolkit of evaluators. However, its presumptions and potentials must fit the research question or evaluation task at hand. We posit that evaluation research will benefit from QCA if the (causal) relationships of interest can be modelled using set theory and formal logic. QCA practitioners should always assume causal complexity when investigating social phenomena as the method rests on three ontological columns: equifinality, conjunctural causation and causal asymmetry (Buche & Siewert, 2015). On that basis, the method expresses social reality in terms of necessary and sufficient conditions for an outcome. Apart from these fundamentals, calibration can build on both quantitative and qualitative raw data, while QCA may easily handle medium numbers of cases in a cross-sectional perspective. In contrast, QCA is probably

not the method of choice if evaluators are interested in linear, additive, correlative and symmetric relationships between variables (Berg-Schlosser et al., 2009, p. 9).

All in all, different methods offer different capability profiles. In this method note, we roughly illuminated the profile of Qualitative Comparative Analysis against the background of evaluation research. A final trait of the set-theoretic approach concerns its flexibility: To unveil causal relationships, QCA may easily be combined with qualitative case studies and process tracing (Schneider & Rohlfing, 2013; Rohlfing & Schneider, 2018) or other methods to triangulate the findings of evaluation studies.

References

Baumgartner, M. & Thiem, A. (2020). Often Trusted but Never (Properly) Tested: Evaluating Qualitative Comparative Analysis. *Sociological Methods & Research*, *49*(2), 279–311. https://doi.org/10.1177/0049124117701487

Befani, B. (2013). Between Complexity and Generalization: Addressing Evaluation Challenges with QCA. *Evaluation, 19(3)*, 269–283. https://doi.org/10.1177/1474022213493839

Berg-Schlosser, D. & Cronqvist, L. (2012). *Aktuelle Methoden der vergleichenden Politikwissenschaft: Einführung in konfigurationelle (QCA) und makro-quantitative Verfahren*. Barbara Budrich.

Berg-Schlosser, D., de Meur, G., Rihoux, B., & Ragin, C.C. (2009). Qualitative Comparative Analysis (QCA) as an Approach. In B. Rihoux & C.C. Ragin (Eds.), *Configurational Comparative Methods. Qualitative Comparative Analysis (QCA) and Related Techniques* (pp. 1–18). Sage.

Blatter, J., Janning, F., & Wagemann, C. (2007). *Qualitative Politikanalyse: Eine Einführung in Forschungsansätze und Methoden*. VS Verlag für Sozialwissenschaften.

Blatter, J., Langer, P.C., & Wagemann, C. (2018). *Qualitative Methoden in der Politikwissenschaft: Eine Einführung*. Springer VS.

Brumm, K. (2022). Schuldenbremse trifft deutschen Föderalismus. Was prägte die Ausgestaltung der neuen Fiskalregeln in den Bundesländern? *Zeitschrift für Vergleichende Politikwissenschaft*, *16*(1), 57–100. https://doi.org/10.1007/s12286-022-00521-z

Brumm, K. & Hörisch, F. (2018). Der lange Schatten der Finanzkrise: Bestimmungsfaktoren fiskalischer Sparpakete in der OECD-Welt. *Zeitschrift für Vergleichende Politikwissenschaft*, *12*(3), 475–507. https://doi.org/10.1007/s12286-018-0396-7

Buche, J. & Siewert, M.B. (2015). Qualitative Comparative Analysis (QCA) in der Soziologie – Perspektiven, Potentiale und Anwendungsbereiche. *Zeitschrift für Soziologie*, *44*(6), 386–406. https://doi.org/10.1515/zfsoz-2015-0602

Duşa, A. (2018). *QCA with R: A Comprehensive Resource*. Springer International Publishing.

Duşa, A. (2022). Critical Tension: Sufficiency and Parsimony in QCA. *Sociological Methods & Research*, *51*(2), 541–565. https://doi.org/10.1177/0049124119882456

Finn, V. (2022). A Qualitative Assessment of QCA: Method Stretching in Large-N Studies and Temporality. *Quality & Quantity, 56(5)*, 3815–3830. https://doi.org/10.1007/s11135-021-01278-5

Freitag, M. & Schlicht, R. (2009). Educational Federalism in Germany: Foundations of Social Inequality in Education. *Governance*, *22*(1), 47–72. https://doi.org/10.1111/j.1468-0491.2008.01421.x

Gerring, J. (2008). Case Selection for Case-Study Analysis: Qualitative and Quantitative Techniques. In J.M. Box-Steffensmeier, H.E. Brady, & D. Collier (Eds.), *The Oxford Handbook of Political Methodology* (pp. 645–684). Oxford University Press.

Gerrits, L. & Verweij, S. (2016). Qualitative Comparative Analysis as a Method for Evaluating Complex Cases: An Overview of Literature and a Stepwise Guide with Empirical Application. *Zeitschrift Für Evaluation*, *15*(1), 7–22. https://www.waxmann.com/artikelART101879

Glatman-Freedman, A., Cohen, M., Nichols, K. A., Porges, R. F., Rayos Saludes, I., Steffens, K., Rodwin, V. G., & Britt, D. W. (2010). Factors Affecting the Introduction of New Vaccines to Poor Nations: A Comparative Study of the Haemophilus influenzae Type B and Hepatitis B Vaccines. *PLoS ONE*, *5*(11), 1–9. https://doi.org/10.1371/journal.pone.0013802

Holvoet, N. & Dewachter, S. (2013). Multiple Paths to Effective National Evaluation Societies: Evidence from 37 Low- and Middle-Income Countries. *American Journal of Evaluation*, *34*(4), 519–544. https://doi.org/10.1177/1098214013497083

Hörisch, F. (2012). Die Bestimmungsfaktoren der Einführung der Unternehmensmitbestimmung in den OECD-Staaten: Eine Fuzzy-Set Qualitative Comparative Analysis. *Zeitschrift für Sozialreform*, *58*(1), 33–57. https://doi.org/10.1515/zsr-2012-0102

Hörisch, F. & Heiken, M. (2020). Qualitative Comparative Analysis: Mit einer Anwendung zur Asylbewerberunterbringung im Bundesländervergleich. In M. Tausendpfund (Ed.), *Fortgeschrittene Analyseverfahren in den Sozialwissenschaften: Ein Überblick* (pp. 243–274). Springer VS.

Jäckle, S. (2015). QCA. In A. Hildebrandt, S. Jäckle, F. Wolf, & A. Heindl (Eds.), *Methodologie, Methoden, Forschungsdesign: Ein Lehrbuch für fortgeschrittene Studierende der Politikwissenschaft* (pp. 191–214). Springer VS.

Lam, W. F. & Ostrom, E. (2010). Analyzing the Dynamic Complexity of Development Interventions: Lessons from an Irrigation Experiment in Nepal. *Policy Sciences*, *43*, 1–25. https://doi.org/10.1007/s11077-009-9082-6

Lauth, H.-J., Pickel, G., & Pickel, S. (2009). *Methoden der vergleichenden Politikwissenschaft. Eine Einführung*. Springer VS.

Mahoney, J. & Goertz, G. (2006). A Tale of Two Cultures: Contrasting Quantitative and Qualitative Research. *Political Analysis*, *14*(3), 227–249. https://doi.org/10.1093/pan/mpj017

Mello, P. A. (2021). *Qualitative Comparative Analysis: An Introduction to Research Design and Application*. Georgetown University Press.

Oana, I.-E., Schneider, C. Q., & Thomann, E. (2021). *Qualitative Comparative Analysis Using R: A Beginner's Guide*. Cambridge University Press.

Ragin, C. C. (1987). *The Comparative Method: Moving Beyond Qualitative and Quantative Strategies*. University of California Press.

Ragin, C. C. (2000). *Fuzzy-Set Social Science*. University of Chicago Press.

Ragin, C. C. (2008). *Redesigning Social Inquiry: Fuzzy Sets and Beyond*. University of Chicago Press.

Ragin, C. C. & Sonnett, J. (2005). Between Complexity and Parsimony: Limited Diversity, Counterfactual Cases, and Comparative Analysis. In S. Kropp & M. Minkenberg (Eds.), *Vergleichen in der Politikwissenschaft* (pp. 180–197). Springer VS.

Rohlfing, I. & Schneider, C. Q. (2018). A Unifying Framework for Causal Analysis in Set-Theoretic Multimethod Research. *Sociological Methods & Research*, *47*(1), 37–63. https://doi.org/10.1177/0049124115626170

Rubinson, C., Gerrits, L., Rutten, R., & Greckhamer, T. (2019, July 22). *Avoiding Common Errors in QCA: A Short Guide for New Practitioners*. COMPASSS. https://compasss.org/wp-content/uploads/2019/07/Common_Errors_in_ QCA.pdf

Rutten, R. (2022). Applying and Assessing Large-N QCA: Causality and Robustness from a Critical Realist Perspective. *Sociological Methods & Research, 51(3)*, 1211–1243. https://doi.org/10.1177/0049124120914955

Schneider, C. Q. (2016). Real Differences and Overlooked Similarities: Set-Methods in Comparative Perspective. *Comparative Political Studies*, *49*(6), 781–792. https://doi.org/10.1177/0010414015626454

Schneider, C. Q. & Rohlfing, I. (2013). Combining QCA and Process Tracing in Set-Theoretic Multi-Method Research. *Sociological Methods & Research, 42*(4), 559–597. https://doi.org/10.1177/0049124113481341

Schneider, C. Q. & Wagemann, C. (2012). *Set-Theoretic Methods for the Social Sciences: A Guide to Qualitative Comparative Analysis*. Cambridge University Press.

Thiem, A. (2014). Navigating the Complexities of Qualitative Comparative Analysis: Case Numbers, Necessity Relations, and Model Ambiguities. *Evaluation Review*, *38*(6), 487–513. https://doi.org/10.1177/0193841x14550863

Thiem, A. (2022). Beyond the Facts: Limited Empirical Diversity and Causal Inference in Qualitative Comparative Analysis. *Sociological Methods & Research, 51*(2), 527–540. https://doi.org/10.1177/0049124119882463

Thomann, E. (2017, September 4). *Why I Still Don't Prefer Parsimonious Solutions* [blog post]. RE-DESIGN. http://re-design.dimiter.eu/?p=915

Vis, B. (2012). The Comparative Advantages of fsQCA and Regression Analysis for Moderately Large-N Analyses. *Sociological Methods & Research, 41*(1), 168–198. https://doi.org/10.1177/0049124112442142

Wagemann, C. (2023). Qualitative Comparative Analysis. In G. Wenzelburger & R. Zohlnhöfer (Eds.), *Handbuch Policy-Forschung* (pp. 397–419). Springer VS.

Prof. Dr. Felix Hörisch | Hochschule der Bundesagentur für Arbeit (HdBA) | Campus Mannheim | Seckenheimer Landstraße 16, D-68163 Mannheim | E-Mail: felix.hoerisch@arbeitsagentur.de

Kai Brumm | Heidelberg University | Institute of Political Science | Bergheimer Straße 58, D-69115 Heidelberg | E-Mail: K.Brumm@stud.uni-heidelberg.de